

# Hlasová komunikácia v inteligentnom automobile (1)

Mikuláš Alexík, Viliam Mateička

Článok opisuje štruktúru programového vybavenia palubného počítača inteligentného automobilu, ktorá zabezpečuje analýzu hlasových príkazov nezávisle od hovoriaceho. Izolované slová príkazov boli overované na špecializovanom signálovom procesore Blackfin BF 561 fy. Analog Devices. Rozpoznávanie príkazov nezávisle od hovoriaceho zabezpečovalo spracovanie na báze skrytých Markovovských modelov. Autori príspevku pokračujú v realizácii verzie vhodnej na komerčné použitie.

## Uvod

Hlasové aplikácie sú dôležité pre ďalší rozvoj komfortných systémov v automobile. Ak bude môcť vodič ovládať jednotlivé aplikácie pomocou hlasových povelov, bude sa môcť lepšie sústrediť na jazdu. Ide o aplikácie, napríklad otváranie a zatváranie okien, ovládanie spätných zrkadiel, zapínanie a vypínanie rádia, ovládanie jeho hlasitosti. V spolupráci s navigačnými systémami možno využiť hlasové aplikácie pri navigovaní vodiča jednoduchými hlasovými povelmi. Nové aplikácie by mali byť schopné prečítať vodičovi SMS správu alebo e-mail. Vodič sa bude môcť venovať ceste a nebude musieť spustiť zrak z vozovky. Celkové zjednodušenie komunikácie medzi človekom a vozidlom zabezpečí až rozpoznávanie reči. Je to proces, pri ktorom počítač identifikuje hovorené slová. To znamená, že pri rozprávaní počítaču dostaneme korektnú odpoveď. Predpokladá to existenciu systému rozpoznávania, ktorý nezávisí od hovoriaceho. Problémom je jazyk a slovná zásoba systému. Čím menšia je slovná zásoba, tým je pre počítač jednoduchšie slová rozoznávať. Systém nie je vždy schopný správne rozoznať dané slovo. V súčasnosti je úspešnosť rozpoznávania pri najlepších systémoch (pre angličtinu) okolo 98 %.

V ďalšom texte je opísaný signálový procesor a rozpoznávanie obmedzeného súboru slov pri komunikácii vodiča v automobile.

## Sentencie pre hlasovú komunikáciu v automobile

Predpokladá sa hlasová komunikácia týmto spôsobom: oznamovanie vodičovi počas navigácie automobilu navigačným systémom; povely sú označené v zozname ako *Italic*; analýza pokynov vodiča a realizácia pokynu; predpokladáme generovanie (*Italic*) a analýzu týchto slov a pokynov:

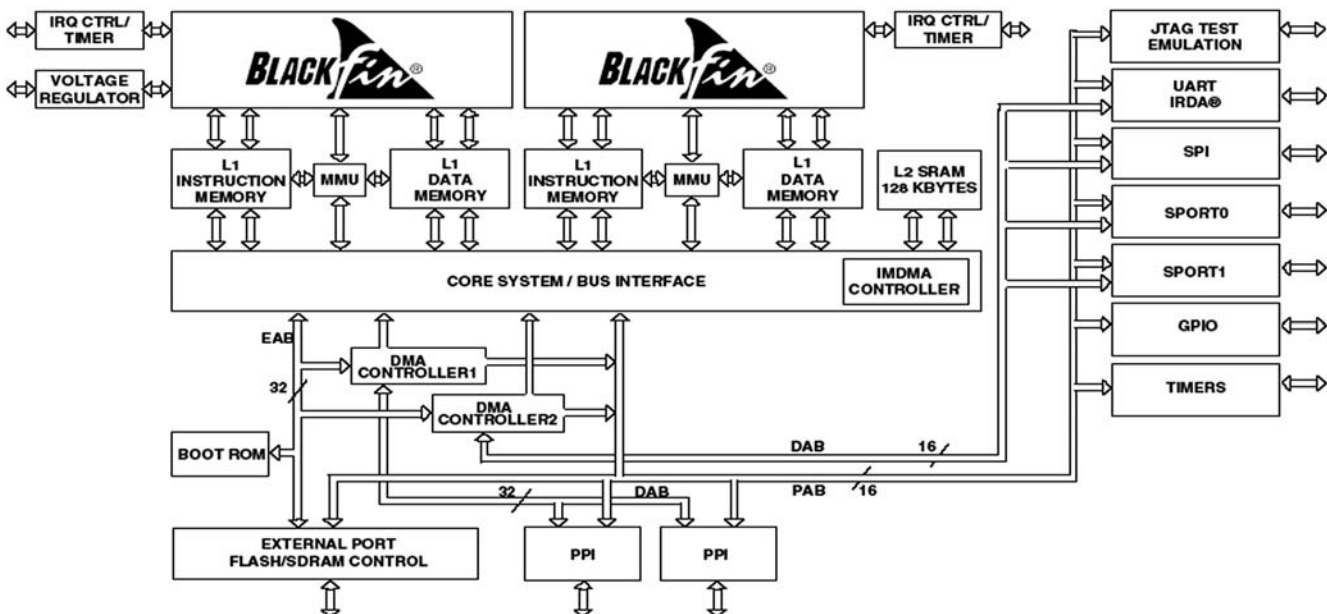
Zatvor	Predné	Okno	Jedna	<i>Prvá</i>
Otvor	Zadné	Rádio	Dva	<i>Druhá</i>
Zapni	Ľavé	Prehrávač	Tri	<i>Tretia</i>
Vypni	Pravé	Displej	Štyri	<i>Štvrtá</i>
Zobraz	Horné	Stanicu	Päť	<i>Piata</i>
Vymaž	<i>Vľavo</i>	<i>Odbočka</i>		<i>Číslo</i>
Odboč	<i>Vpravo</i>	<i>Pokračuj</i>		<i>Zastav</i>

Z uvedených slov možno skladať oznámenia palubného počítača počas navigácie: odboč, tretia odbočka vpravo, pokračuj, tretia odbočka vľavo a pod.

Modul pre hlasovú komunikáciu by mal analyzovať vybrané slová a vykonať patričný úkon po zadaní týchto príkazov: zapni prehrávač, otvor predné okno pravé, zapni stanicu číslo tri a pod. Kombináciou slov vznikajú zmysluplné príkazy. Poradie slov v príkaze je jednoznačne dané preto, aby sa zjednodušila analýza vety procesorom. Kombinácie slov, resp. použiteľné vety pre modul rozpoznávania príkazov nezávisí od rečníka a netreba ho podrobne rozpisovať. Kľúčové je rozpoznávanie 22 uvedených slov.

## Technický prostriedok na overovanie hlasovej komunikácie

Ako technický prostriedok bol aplikovaný vývojový modul signálového procesora Blackfin BF561 od firmy Analog Devices na vývojovej doske EZ-KIT. Obsahuje tieto interné a externé zariadenia integrované na vývojovej doske:



Obr.1 Bloková schéma procesora Blackfin BF-561



- dve nezávislé jadrá procesorov, špecifikované ďalej,
- L1 pamäť na inštrukcie, dáta a medzivýsledky, nezávisle pre obe procesorové jadrá, konfigurovateľná aj ako cache pamäť (2 x 16 M + 4 x 16 M + 4 M),
- L2 spol. pamäť druhej úrovne L1 a L2 – v puzdre procesora,
- dvojvrstvový prerušovací podsystem (2x radiče prerušenia jadra, 1x radič systému),
- 64 MB RAM externej pamäte (16 M x 16 bit x 2),
- 8 MB FLASH externej pamäte (4 M x 16 bit),
- viacanálový audiokodek (AD1836), ktorý obsahuje 24-bitový 96 kHz A/D a D/A prevodníky, 8 vstupov a 12 výstupov,
- rýchly trojkanálový 10-bitový A/D prevodník (ADV7183A),
- trojkanálový D/A prevodník (ADV7179) na kódovanie videesignálu vo formáte NTSC/ PAL,
- dva nezávislé 16-kanálové DMA kontroléry,
- 14 program. časovačov, z nich 12 na prerušenie procesora,
- dva UA sériové porty, asynchrónny sériový port UART,
- dve univerzálne programovateľné paralelné rozhrania,
- 48 obojsmerných I/O pinov na všeobecné použitie, v realizovanej aplikácii sú použité pre 4 tlačidlá, 16 LED diód a A/D aj D/A prevodníky.

Najdôležitejšou procesora sú dve nezávislé Blackfin procesorové jadrá. Celkový duálny symetrický výkon možno regulovať nastavením pracovnej frekvencie jadra až do 600 MHz. Blackfin procesor kombinuje zmiešanú 16- a 32-bitovú sadu inštrukcií podobnú RISC, navrhnutú pre optimálnu hustotu kódu a umožňuje až 10-stupňové prúdové spracovanie týchto inštrukcií. Každé jadro obsahuje dve 16-bitové násobičky (MAC), dve 32-bitové ALU, štyri 8-bitové video ALU a jeden posuvný 40-bit register. Tieto jednotky spracúvajú 8-, 16- alebo 32-bitové dáta. Pri paralelnom spracovaní je jedno jadro schopné vykonať až štvoricu 16-bitových operácií v každom cykle. 40-bitový posuvný register je schopný vykonávať bitovú rotáciu, posun, normalizáciu, extrakciu a dekompozíciu.

K vývojovému prostrediu je k dispozícii proprietárny operačný systém (OS) Visual DSP Kernel (VDK), poskytujúci jednoduchý OS s multitaskingom, ktorý nebolo potrebné v tejto aplikácii použiť.

K dispozícii boli dve procesorové jadrá fungujúce paralelne, prvé bolo použité na predspracovanie signálu a extrakciu vlastností, druhé pri implementácii HMM (počítanie výstupnej pravdepodobnosti jednotlivých pozorovaní a vyhľadávanie v jednotlivých modeloch Viterbiho algoritmom), obr. 3 (pozn. redakcie: obr. 3 bude publikovaný v druhej časti tohto článku).

Aplikácia systému rozpoznávania izolovaných slov je realizovaná na doske (BF561 EZ-KIT Lite) k procesoru Blackfin-BF561 a vytvorená vo vývojovom prostredí VisualDSP++ s použitím programovacieho jazyka C.

### Špecifikácia aplikovaného systému rozpoznávania reči

Realizovaná implementácia systému rozpoznávania má tieto vlastnosti: rozpoznáva iba izolované slová, nezávisí od rečníka (miera tejto nezávislosti je daná počtom rečníkov použitých na natrénovanie jednotlivých slov), malý slovník obsahuje 32 slov, model jazyka je z konečným počtom stavov, zložitost je malá až stredná, odstup signál – šum (SNR) je dostatočne veľký pri použití dynamického mikrofónu s citlivosťou 70 dB, nízkošumového predzosilňovača a 16-bitového A/D prevodníka (použitím SNR prevodníka je 98 dB). Zdrojom signálu je dynamický vokálový mikrofón, ktorý je použitý na proces tréningu aj rozpoznávania. Ako referenčný program na implementáciu aplikácie rozpoznávania reči v procesore ADSP-Blackfin BF-561 a ako nástroj na vytvorenie modelov jednotlivých slov bol použitý opensource balík

Hidden Markov Model Toolkit – HTK Book [http://htk.eng.cam.ac.uk/] a jeho knižnica HTKLib [1]. Pre aplikáciu sme sa rozhodli použiť HMM hlavne z dôvodu jednoduchého tréningu modelov a možnosti ovplyvňovať jednoduchými zmenami parametrov modelov celkový výkon rozpoznávača reči.

Postup rozpoznávania z obr. 2 možno rozdeliť na bloky:

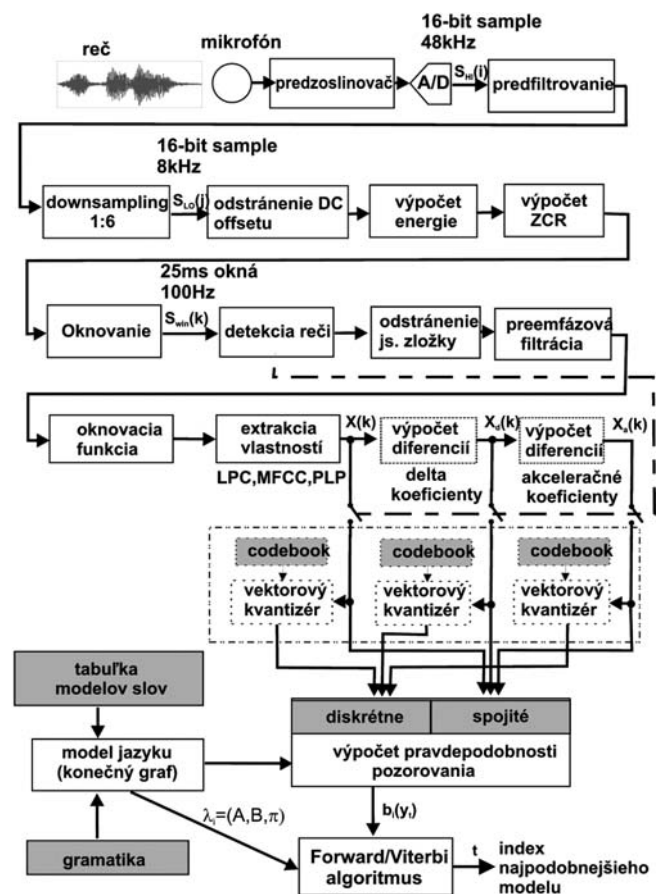
- úprava a predspracovanie signálu,
- extrakcia akustických vlastností reči, modelovanie a klasifikácia slov.

### Úprava a predspracovanie signálu

Úpravou signálu sa rozumie predfiltrovanie, downsampling, odstránenie offsetu a výpočet energie a ZCR. Predspracovanie (oknovanie) znamená detekciu reči, odstránenie jednosmernej zložky a preemfázovú filtráciu.

Získaný analógový signál z mikrofónu je zosilnený a filtrovaný v predzosilňovači, vzorkovaný a kvantovaný na diskretný signál v A/D prevodníku. Ľudská reč má rozsah frekvencií 300 Hz až 3,4 kHz. Použitý A/D prevodník nemožno použiť na nižšej vzorkovacej frekvencii ako 48 kHz, preto treba tento signál prevzorkovať na nižšiu frekvenciu (8 kHz), kde je už vhodný na analýzu rečového signálu. Pred vzorkovaním treba odfiltrovať vyššie frekvencie, aby neovplyvňovali signál. Na prevzorkovanie bola aplikovaná jednoduchá aproximácia aritmetickým priemerom, ktorá sa správa ako FIR filter 5. rádu.

**Oknovanie.** Vykonáva rozdelenie navzorkovaných amplitúd signálu do prekrývajúcich sa tzv. okien, v ktorých ďalej prebieha spracovanie. Pri oknovaní bol signál so vzorkovacou frekvenciou 8 kHz (interval 125 us) rozdelený do 25 ms okien (po 200 vzoriek) s prekrývaním 15 ms (120 vzoriek), posun okna je potom 10 ms (80 vzoriek). S touto periódou už ďalej pracuje celý systém rozpoznávača, preto je jej voľba kritická pre celý návrh. Každé okno reprezentuje vlastnosti reči v intervale 15 až 30 ms dlhom.



Obr.2 Bloková schéma systému rozpoznávania reči s HMM



**Detekcia reči.** Ide o zistenie prítomnosti užitočného signálu – signálu reči, resp. detegovanie začiatku a konca slova. Aplikovaná bola kombinácia tzv. „zero-crossing rate“ a výpočtu krátkodobej energie. V tomto bloku je aplikované aj potlačenie šumu pozadia.

**Odstránenie jednosmernej zložky.** Prebieha v jednotlivých oknách. Táto časť spracovania nemusí byť implementovaná, ak je použitá preemfázová filtrácia. Preemfáza v podstate predstavuje FIR filter prvého rádu. Jeho koeficient  $k$  sa v systémoch rozpoznávania reči obyčajne volí  $k = 0,97$ , čo zabezpečí potlačenie signálov s malou frekvenciou a jednosmernej zložky.

**Oknovacia funkcia.** Aplikovanie oknovacej funkcie vyhladí nespojitosti vzniknuté rozdeľovaním do jednotlivých okien. Najpoužívanejšou oknovacou funkciou pri spracovaní rečového signálu je Hammingova funkcia. Touto oknovacou funkciou potom vynásobíme vstupný signál. V DSP je táto funkcia na zníženie výpočtovej zložitosti predpočítaná iba raz a uložená do poľa.

### Literatúra

[1] MATEIČKA, V.: Implementácia systému rozpoznávania reči nezávislého od rečníka do DSP Blackfin-BF561. Diplomová práca, FRI, KTK ŽU v Žiline, 2006

[2] HOLADA, M.: (1999) Comparing diferent parameterization methods for telephone speech re-cognition. In: Proc. of 9<sup>th</sup> Czech-German Workshop Speech Processing, Prague, September 1999.

[3] Survey of the State of the Art in Human Language Technology 1996 <http://cslu.cse.ogi.edu/HLTSurvey/index.html>

[4] Hidden Markov Toolkit – HTKBook <http://htk.eng.cam.ac.uk/>

*Pokračovanie v budúcom čísle.*

**prof. Ing. Mikuláš Alexík, PhD.  
Viliam Mateička**

**Žilinská univerzita  
Fakulta riadenia a informatiky  
Katedra technickej kybernetiky  
Veľký diel, 010 26 Žilina  
e-mail: Mikulas.Alexik@fri.utc.sk**

46